

TECHNICAL PAPER

RAC-T-166

JULY 1965

D623094

RESEARCH ANALYSIS CORPORATION

Optimal Decision Rules for the E Model of Chance-Constrained Programming

CLEARINGHOUSE
FOR FEDERAL SCIENCE AND
TECHNICAL INFORMATION

Approved for public release
by the Defense Science Board
Michael J. L. Kirby
Research Analyst, Corporation



PRINTED FOR THE
DEPARTMENT OF THE ARMY
Contract No. DA-34-186-3-1

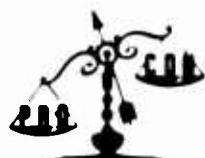
Copy 73 of 750

ADVANCED RESEARCH DIVISION
TECHNICAL PAPER RAC-TP-166
Published July 1965

Optimal Decision Rules for the E Model of Chance-Constrained Programming

by
Abraham Charnes
Northwestern University

Michael J. L. Kirby
Research Analysis Corporation



RESEARCH ANALYSIS CORPORATION

MCLEAN, VIRGINIA

FOREWORD

This is the first in a series of papers dealing with the problem of finding the optimal decision rules for n -period chance-constrained programming models. The significant feature of this paper is that the admissible class of decision rules is not required to possess any specific analytic property, as it is in all previous work in this field. Instead the admissible class is the largest possible class of decision rules consistent with the interpretation given to n -period problems. It is shown that in this case the optimal decision rules are piecewise linear functions of certain conditional fractile points and the decision rules of all preceding periods.

Mr. Charnes' work was supported in part by the Office of Naval Research [Research Contract Nonr-1228 (10), Project 047-021] and by the National Institutes of Health (Projects EF 00355-01 and WP 00019-04). Mr. Kirby's work has been conducted in the research program of RAC's Advanced Research Division, under Department of Army sponsorship.

Nicholas M. Smith
Chief, Advanced Research Division

CONTENTS

Foreword	iii
Abstract	2
1. Introduction	3
2. A Chance-Constrained Example: The Stochastic Heating-Oil Problem	4
3. Some Choices of Objectives in Chance-Constrained Programming	7
4. Choice of Decision Rules	9
5. Multiperiod Chance-Constrained Problems	11
Theorem 1	
6. The General n-Period E Model	12
7. Matrix Theory	17
8. Remarks about the Consistency of the Constraints	18
9. Optimal Class of Decision Rules	21
Theorem 2	
10. Optimal Decision Rules for Some Special Cases	24
Theorem 3	
11. Examples	26
Example 1—Example 2	
References	33
References Cited—Additional References	

**Optimal Decision Rules
for the E Model
of Chance-Constrained Programming**

ABSTRACT

The first five sections of this paper contain an introduction to the topic of chance-constrained programming. Then the general n -period expectation-objective model of chance-constrained programming is presented and certain necessary conditions are established for decision rules to be optimal for such a model. The question of the consistency of the constraints and the finiteness of the optimal value of the objective function for such problems is discussed and several methods of resolving these questions are presented. The simplification that results when the chance-constrained problem is treated as a problem of linear programming under uncertainty is also discussed. The paper is concluded by solving two two-stage problems.

1. INTRODUCTION

As an introduction to the topic of chance-constrained programming, the various kinds of objective functions and admissible classes of decision rules for chance-constrained problems that have appeared in the literature are discussed in this paper, and remarks are illustrated by using the stochastic heating-oil problem. The general n -period expectation-objective model is then presented and optimal decision rules for such a model are shown to be piecewise linear functions of certain conditional fractile points. Using these results, the simplification obtained by treating the problem as one of linear programming under uncertainty is discussed. The paper is concluded by solving two rather simple two-stage problems.

The topic of chance-constrained programming is, perhaps, best introduced by first examining an ordinary linear-programming problem in its most general form, i.e.,

maximize

$$c^T X$$

subject to

$$AX \leq b \quad (1)$$

where A is an $m \times n$ matrix of constants, and b, c are, respectively, $m \times 1$ and $n \times 1$ constant vectors. The object of problem 1 is to find the $n \times 1$ vector X from the set of all X satisfying $AX \leq b$, which maximizes $c^T X$.

If a chance-constrained formulation were to replace problem 1 a problem of the following type would be obtained:

maximize

$$f(c, X)$$

subject to

$$P(AX \leq b) \geq \alpha, \quad (2)$$

where P means probability. Here A, b, c are no longer constant matrices as they were in problem 1; rather, they can have some or all of their elements as random variables. The $m \times 1$ vector α contains a prescribed set of constants that are probability measures of the extent to which constraint violations are permitted. In other words, an element $\alpha_i \in \alpha$ and satisfying $0 \leq \alpha_i < 1$ is associated with the constraint

$$\sum_{j=1}^n a_{ij} x_j \leq b_i$$

to give

$$P\left(\sum_{j=1}^n a_{ij} x_j \leq b_i\right) \geq \alpha_i \quad (3)$$

This means that the i th constraint may on occasion be violated, but it can be violated, at most, $100(1-\alpha_i)$ percent of the time.

As yet nothing has been said about the nature of the functional of problem 2. As might be expected, a fairly wide range of reasonable choices can be considered for the function $f(c, X)$. In Section 3 some of these choices will be discussed and the work that has been done on them will be summarized.

An important question in chance-constrained programming concerns the classes of decision rules that should be admissible for the problem. Because the object in problem 2 is to find an optimal vector of stochastic decision rules $X = \varphi(A, b, c)$ with the function φ chosen from some prescribed class of functions, a fundamental question is how to choose the admissible class of functions for φ . In all the literature published thus far, it is assumed that φ is a linear function of the elements of b , i.e., X is restricted to a linear decision rule. In the results of this paper the admissible class of decision rules is greatly enlarged and then the optimal decision rules are shown to be piecewise linear. This result is of particular importance because it shows that there exists good reason, other than the fact that it is mathematically more manageable, to limit oneself to finding the optimal linear rule.

In some cases the class from which φ is to be chosen may be restricted deliberately, so that at the time of application of the decision rules X is not a function of some random variables; rather, the resulting X values are assigned numerical, i.e., deterministic, values. When this can be done it is said that the resulting function provides a set of certainty-equivalent relations because it specifies the decision exactly, i.e., with certainty.

In general, however, we permit a choice of φ for which the resulting X vectors are also random. Thus problem 2 is solved for a set of stochastic decision rules imputing action for each manifestation of the random variables involved in the problem. This is in contrast to other possible solution notions such as mixed strategies in game theory or Manne's problem¹—which, it might be noted, corresponds to the solution of only the first problem of the two into which the chance-constrained problem was factored in Charnes and Cooper²—or Tintner's stochastic programming.³

In either the stochastic or certainty equivalent case it may also be possible to develop a deterministic equivalent for problem 2. That is, it may be possible to find a problem that is equivalent and yet does not contain any of the random variables explicitly. Hence it is a deterministic (often convex) programming problem.

These deterministic equivalents, when attainable, often provide a means of convenient analysis and characterization of the solutions of the given chance-constrained problem. In addition they may also be valuable in their own right for purposes of theoretical analysis and policy review. How to obtain deterministic equivalents for certain choices of $f(c, X)$ when the admissible class of functions for φ is the class of linear functions of b is illustrated in Section 3.

2. A CHANCE-CONSTRAINED EXAMPLE: THE STOCHASTIC HEATING-OIL PROBLEM

In order to elaborate further on the ideas developed above, an example of a chance-constrained problem will be given by using a simplified version of

the problem that first gave rise to the chance-constrained programming formulation and application.⁴ The problem is one of multiperiod scheduling of heating-oil production to meet an uncertain demand that depends heavily on the weather. In the original problem the constraints included marketing constraints, which required production to be planned in such a way that customer demands were met as they materialized, and storage constraints, which restricted the amount of oil that could be kept in inventory. The objective of the problem was to maximize expected profits over the entire planning horizon.

As Charnes, Cooper, and Symonds showed,⁵ a deeper consideration of the constraints suggested that the problem was, in fact, one of minimization of expected costs rather than maximization of the expected profits. Because the constraints of the problem were such that customer demand was to be taken as given, i.e., stochastically determined, the real objective was to supply whatever demands arose at a minimum total expected cost.

A simplified version of this problem is:

minimize

$$E \left[\sum_{j=1}^n (c_j R_j + k_j I_j) \right]$$

subject to

$$\begin{aligned} P(I_0 + \sum_{j=1}^l R_j \geq \sum_{j=1}^l S_j) &\geq \alpha_l \quad l = 1, \dots, n \\ P(R_j \geq 0) &\geq 1 \quad j = 1, \dots, n, \end{aligned} \tag{4}$$

where the symbol E refers to the expectation operator, so the objective is expected value minimization over the n periods of the planning horizon. The c_j and k_j represent, respectively, the refinery costs and the inventory carrying charges for period j ; both c_j and k_j , $j = 1, \dots, n$, may be random variables. I_j is the initial inventory at the start of period j . The random variable S_j represents the anticipated sales for period j . R_j represents the production rate to be scheduled in period j and is the quantity for which a decision rule is to be found.

Thus problem 4 shows that minimization of the total expected production costs over periods $j = 1, \dots, n$ is desired, subject to meeting customer demand at least 100 α_l percent of the time in each period.

The probability α_l is known as the "risk level" associated with the l th constraint. The quantity

$$b_l = \sum_{j=1}^l S_j$$

is called the "quality level." Although in problem 4 these quality levels include all sales, it is important to note that this need not always be the case. For example, we may adjust the original random variables so that b_l refers to a certain fraction of the original sales variables. This would not change the problem in any mathematically essential way but it would alter our interpretation of the problem. Thus the constraints could be changed to mean that in period l , 100 β_l percent of all sales demands, plus or minus a safety margin, must be met at least 100 α_l percent of the time. This flexibility in interpreta-

tion of mathematically similar models is one of the significant features of chance-constrained programming. In solving problem 4 it is convenient to find decision rules such that the actual refinery production rates will be known exactly at the start of the period to which they apply. In other words, it is desired that the stochastic decision rules R_j , $j = 1, \dots, n$, be such that R_j is a known number at the start of the j th period. This means that the admissible class of decision rules for R_j can involve random variables only insofar as their values will be known, i.e., will have been observed, at the time of application of R_j . Hence R_j can be an explicit function of only the random variables of periods $1, \dots, j - 1$, and it cannot be a function of the random variables of the j th or future periods.

In particular, if consideration is limited to linear rules, the rules will be of the form $R_j = \Gamma S_j + \delta_j$, where Γ is a lower triangular matrix. Specifically, we can write

$$R_j = \sum_{l=1}^{j-1} \gamma_{jl} S_l + \delta_j \quad l = 1, \dots, n. \quad (5)$$

where each γ_{jl} , δ_j is a scalar and $\gamma_{10} = 0 = \gamma_{11}$.

In their work, Charnes, Cooper, and Symonds⁴ were able, by using Eq 5, to convert problem 4 into a deterministic equivalent, which was then solved to give the optimal values γ_{jl}^* and δ_j^* for γ_{jl} and δ_j , respectively. Using these values, the optimal decision rules R_j^* , $j = 1, \dots, n$, for problem 4 are given by

$$R_1^* = \delta_1^*.$$

$$R_2^* = \gamma_{21}^* S_1 + \delta_2^*.$$

$$R_3^* = \gamma_{31}^* S_1 + \gamma_{32}^* S_2 + \delta_3^*.$$

and, in general,

$$R_j^* = \sum_{k=1}^{j-1} \gamma_{jk}^* S_k + \delta_j^*.$$

It is important to observe that different γ_{jk}^* weights apply to the same observed value of S_k . Thus S_1 need not (and in general will not) receive the same weight in R_j^* as it did in R_2^* . But once γ_{jk}^* , δ_j^* , $j > k$, $j = 1, \dots, n$, have been found, Eq 5 supplies a set of relations (i.e., decision rules) that can be used to generate the required numbers (i.e., decisions) at the time these numbers are needed. Thus, as the observations on the preceding S_j values are obtained, the resulting R_j values will be known with the certainty the problem requires.

The certainty-equivalent relations discussed above may supply all that is required for the conduct of operations. They may not, however, meet all the needs of management, since management may want to evaluate different alternatives before committing itself to a given set of policies or actions. Since Eq 5 cannot be completed until the S_j values have been observed, this evaluation of different alternatives cannot be achieved. However, when a deterministic equivalent that assumes the form of a linear or convex programming problem is available, the duality relations of such problems can be used to supply dual evaluators. Then, prior to obtaining the data needed to achieve the certainty-equivalent relations, it is possible to study the overall effects of variations in

risk levels, quality levels, etc., and in other forms of constraint alteration and data testing.

This completes the discussion of the stochastic heating-oil problem. It has been examined at some length in order to illustrate one possible area of application of the results developed in Sections 6-10 of this paper and in Ref 5. However, it is emphasized that this is not the only area to which chance-constrained programming has been successfully applied. Charnes, Cooper, Devoe, and Learner⁶ consider a problem of selecting studies and statistical estimators in new-product marketing studies. Chance-constrained programming is also very useful in problems of financial budgeting and planning for liquidity. This is illustrated by Charnes and Thore⁷ when they consider the problem of planning for liquidity in a savings and loan association. Their results are extended further in a RAC paper now in preparation.⁸ Also, the investigation of the uses of chance-constrained programming in critical path analysis was begun recently by Charnes, Cooper, and Thompson.⁹

3. SOME CHOICES OF OBJECTIVES IN CHANCE-CONSTRAINED PROGRAMMING

As mentioned in the "Introduction" section, there are many possible choices for the function $f(c, X)$ in problem 2. In fact, it must be emphasized that the very concept of optimization under risk and uncertainty immediately raises important questions about the choice of rational objectives. Such questions arise, for example, concerning the reasonableness of an expected-value optimization. Why not choose some other measure of value? The importance of this question becomes clear when it is noted that the decision rules that are optimal for one problem and a given objective will not, in general, be optimal for the same problem under a different objective.

Charnes and Cooper^{10,11} investigate three different classes of objectives; specifically, they examine (a) an expected-value optimization, (b) a minimum-variance (or mean-square error) objective, and (c) a maximum-probability model. These models are called, respectively, the "E model," "V model," and "P model" of chance-constrained programming.

It is of interest to distinguish between the first two models and the third by reference to what H. A. Simon¹² calls the "satisficing," as opposed to the "optimizing," objective.[†] Simon originally proposed this objective as an alternative to (a) and (b) in order to try to resolve some of the inadequacies of optimizing objectives for characterizing certain aspects of human behavior. In particular, he suggested that human beings do not always seek an absolute extremum before taking action in a given situation; rather, they try to maximize the probability of being better off than some given point of reference.

In the P model, vectors c^0 , X^0 , are specified relative to some set of values that an organism or human being will regard as satisfactory whenever these levels are achieved. Of course the organism confronted by an environment subject to risk cannot be certain of achieving the given level c^0 , X^0 when effecting its choice from what it believes to be the available alternatives. Therefore

[†]See also March and Simon,¹³ and the extensions and applications to problems of budgetary management given in Cooper and Savvas,¹⁴ Stedry,¹⁵ and Stedry and Charnes.¹⁶

it tries to maximize the probability of obtaining at least its desired level $c^0 T X^0$ subject to its feasible alternatives of action.

In their papers, Charnes and Cooper^{10,11} express the E model as follows:

maximize

$$E(c^T X)$$

subject to

$$P(AX \leq b) \geq a, \quad (6)$$

where A is a matrix of constants and b, c are vectors of random variables. They solve problem 6 for the optimal linear rule given by $X = Db$, and so convert the problem to one of finding the elements of the matrix D. Under suitable assumptions concerning the symmetry, and the existence of second moments, of certain distributions† they show that the deterministic equivalent of problem 6 is

minimize

$$-\mu_c^T D \mu_b$$

subject to

$$\begin{aligned} -a_i^T D \mu_b - v_i &\geq -\mu_{b_i}, \quad i = 1, \dots, m, \\ -K_{a_i}^2 E[\hat{b}_i - a_i^T D \hat{b}]^2 + v_i^2 &\geq 0, \quad i = 1, \dots, m, \\ v_i &\geq 0, \quad i = 1, \dots, m, \end{aligned} \quad (7)$$

where $\mu_c^T = [E(c)]^T$, $\mu_b^T = [E(b)]^T$, a_i^T is the i th row of A, $\hat{b}_i = b - \mu_b$, and $-K_{a_i} \equiv F_i^{-1}(\alpha_i)$ where $F_i(\cdot)$ is the cumulative distribution function of the random variable $z_i \equiv (\hat{b}_i - a_i^T D \hat{b}) / E[\hat{b}_i - a_i^T D \hat{b}]^2$. This is a deterministic convex programming problem in the variables v_i , $i = 1, \dots, m$, and the elements d_{ij} of D.

Using the same kinds of techniques and limiting themselves to finding linear decision rules, Charnes and Cooper^{10,11} found deterministic equivalents for the V model:

minimize

$$E(c^T X - c^0 T X^0)^2$$

subject to

$$P(AX \leq b) \geq a,$$

$$X = Db,$$

where $c^0 T$, X^0 are given vectors; and for the P model:

maximize

$$P(c^T X \geq c^0 T X^0)$$

subject to

$$P(AX \leq b) \geq a,$$

$$X = Db.$$

†For similar developments for other kinds of distributions and where D is specialized in various ways see Ben-Israel¹⁷ and Ben-Israel and Charnes.¹⁸

In both these models the resulting deterministic equivalents contain the constraints of problem 7 as part of their system of constraints, although they also have additional constraints and a different objective function.

In the problem to be considered, starting in Section 6, the expectation-objective formulation will be used. No answer is attempted to the question of whether it is the "right" objective to use, but it seems to be a "reasonable" objective for problems of planning over an n -period horizon.

4. CHOICE OF DECISION RULES

The developments of the following sections are concerned with establishing certain properties of the optimal decision rules for the general n -period E-objective model. In Section 6, the admissible class of decision rules for the problem is defined as the most general class of rules consistent with this interpretation of n -period problems. However, in order to illustrate clearly how closely the choice of an admissible class is related to this interpretation of the constraints of the problem, problem 4 must be considered again. There it was seen that a choice of the admissible class to be $R = \Gamma S + \delta$, where Γ is lower triangular, led to a constraint interpretation of the following kind:

$$P(I_\ell + R_\ell \geq S_\ell | I_0 + \sum_{j=1}^{\ell-1} R_j - \sum_{j=1}^{\ell-1} S_j = I_\ell) \geq \alpha_\ell, \ell = 1, \dots, n, \quad (8)$$

that is, given the observations S_j , $j = 1, \dots, \ell-1$, and decisions R_j , $j = 1, \dots, \ell-1$, on the right of the vertical stroke in Eq 8 (the stroke is used to suggest conditional probability), and, before S_ℓ has been observed, a value must be found for R_ℓ that, when added to the ℓ 'th period's beginning inventory I_ℓ , will be sufficient to meet the unknown demand S_ℓ with at least the specified probability α_ℓ .

Clearly then other choices of the matrix Γ will lead to (or be associated with) different ways of interpreting the constraints. This would be true if, for example, R_ℓ were permitted to be an explicit function of S_ℓ .

This problem can be illuminated further by considering some of the work done by Ben-Israel¹⁷ on what will be called the "zero-order decision rule." Here the decision maker wants to know all his program values in advance, i.e., at the start of the planning period. Such a situation can arise, for example, in some aspects of short-term country development planning as well as in certain kinds of budgetary planning practice.

For this particular case Ben-Israel's theorem¹⁷ is employed to obtain the result

$$P\left(\sum_{j=1}^n a_{ij} x_j \leq b_i\right) \geq \alpha_i$$

if and only if

$$\sum_{j=1}^n a_{ij} x_j \leq F_i^{-1}(1-\alpha_i),$$

where $F_i(\cdot)$ is the cumulative marginal distribution function of the random variable b_i and y is identical with $F_i^{-1}(1 - \alpha_i)$ if and only if y equals maximum $\{y : F_i(y) \leq 1 - \alpha_i\}$. Thus the chance-constrained problem

maximize

$$E(c^T x)$$

subject to

$$\begin{aligned} P(AX \leq b) &\geq \alpha, \\ X &\geq 0 \end{aligned}$$

has as its deterministic equivalent

maximize

$$\mu_c^T x$$

subject to

$$\begin{aligned} AX &\leq F^{-1}(1-\alpha), \\ X &\geq 0, \end{aligned} \tag{9}$$

where $\mu_c^T \equiv E(c^T)$.

In a similar manner the vector $G^{-1}(\beta)$ can be defined as such that

$$P(\omega^T A \geq c^T) \geq \beta$$

if and only if

$$\omega^T A \geq [G^{-1}(\beta)]^T$$

Thus if $G_j(\cdot)$ is the cumulative marginal distribution function of c_j , the j -th component of c^T , then $y \equiv G_j^{-1}(\beta)$ if and only if $y = \min\{y : G_j(y) > \beta\}$.

Therefore the deterministic equivalent of

minimize

$$E(\omega^T b)$$

subject to

$$\begin{aligned} P(\omega^T A \geq c^T) &\geq \beta, \\ \omega^T &\geq 0 \end{aligned}$$

is: minimize

$$\omega^T \mu_b$$

subject to

$$\begin{aligned} \omega^T A &\geq [G^{-1}(\beta)]^T, \\ \omega &\geq 0, \end{aligned}$$

where $\mu_b \equiv E(b)$.

The result obtained is the same as that of Ben-Israel, i.e., each of the following two problems is the dual of the other:

maximize

$$[G^{-1}(\beta)]^T x$$

subject to

$$P(AX \leq b) \geq \alpha.$$

$$x \geq 0$$

and minimize

$$\omega^T F^{-1}(1-\alpha)$$

subject to

$$P(\omega^T A \geq c^T) \geq \beta.$$

$$\omega^T \geq 0.$$

This example clearly illustrates that the choice of decision rules here (or, alternatively, the interpretation of the constraints) markedly affects the resulting mathematical treatment of the model. For example, in the case of zero-order decision rules it is possible to develop a duality theory along the lines of the foregoing work, making the solution and evaluation of such problems much simpler than when much larger classes of decision rules are considered.

It would be possible to extend this discussion further by investigating other special and interesting cases. Instead, however, n -period problems will be considered again and interpreted. In the following developments the interpretation first suggested in Section 2 will be used. That is, the admissible class of rules for the j th-period decision rule can include explicitly only the random variables whose values will have been observed when the time comes to put the j th decision into effect. Thus, as in the heating-oil example, the j th-period decision rule will be a known number at the start of the j th period. However, unlike the developments outlined in Section 3, no other restrictions will be placed on the rules. Hence a much larger class of rules than the set of linear rules already discussed will be considered, and yet the admissible class will include the class of linear rules.

5. MULTIPERIOD CHANCE-CONSTRAINED PROBLEMS

Before proceeding with the solution of the general n -period β -objective problem, a brief survey will be presented of the literature on n -period problems as it pertains to this discussion. As has been mentioned, the work of Charnes, Cooper, and Symonds⁴ is fundamental in the field. In addition the work of Symonds¹⁹ on stochastic scheduling by the horizon method and Charnes and Thor⁷ on liquidity planning are mentioned.

If the special case of chance-constrained programming, which is obtained by making all the constraints hold with probability one, is considered, the problem obtained is one known in the field as "linear programming under uncertainty." Such problems were first discussed by Dantzig²⁰ in relation to what he called the " n -stage problem." This emphasis on staging does not, however, provide a wholly adequate characterization of the problem, for—as was shown by Charnes, Cooper, and Thompson²¹—the n -stage problem of Dantzig can be converted into

a one-stage ordinary minimization problem. Thus, whereas the staging emphasis may appear to have a certain appeal in distinguishing between this and other approaches to programming, when statistical errors are present in the matrices (as, for example, in Madansky^{22, 23}), it also tends to conceal the simpler characterizations of the problem.

Specifically, Charnes, Cooper, and Thompson²¹ established the following important result:

Theorem 1

Consider the problem

maximize

$$E[f(X_1, \dots, X_m; b_1, \dots, b_m)]$$

subject to

$$\sum_{j=1}^l X_j = b_i, \quad i = 1, \dots, m$$

$$X_j \geq 0$$

where the X_j are piecewise analytic functions of the b_j and X_i , and f is a linear function of X_1, \dots, X_n . Then the optimal decision rule for X_i is a piecewise linear function of b_1, \dots, b_l and X_i .

In the following work a similar result will be obtained as an immediate corollary to the main theorem.

6. THE GENERAL n -PERIOD E MODEL

The problem to be discussed is the following:

maximize

$$E(c_1^T X_1 + c_2^T X_2 + c_3^T X_3 + \dots + c_n^T X_n)$$

subject to

$$\begin{aligned} P(b_1 \leq A_{11} X_1) &\leq a_1 \\ P(b_2 \leq A_{21} X_1 + A_{22} X_2) &\leq a_2 \\ P(b_3 \leq A_{31} X_1 + A_{32} X_2 + A_{33} X_3) &\leq a_3 \\ &\vdots \\ P(b_n \leq A_{n1} X_1 + A_{n2} X_2 + A_{n3} X_3 + \dots + A_{nn} X_n) &\leq a_n \\ X_j &\geq 0 \quad j = 1, \dots, n \end{aligned} \tag{10}$$

where P stands for probability and E denotes the expected-value operator. (A more detailed explanation of these follows.)

In problem 10 it is assumed that—

- (a) A_{ij} , $i \geq j, i, j = 1, \dots, n$, is an $m_i \times n_j$ matrix all the elements of which are known constants.
- (b) α_i , $i = 1, \dots, n$, is an $m_i \times 1$ vector in which each element is a prescribed probability, i.e., $\alpha_i^T = (\alpha_{i1}, \dots, \alpha_{im_i})$ is such that $0 \leq \alpha_{ik} \leq 1$ for $k = 1, \dots, m_i$, $i = 1, \dots, n$.
- (c) $c_i^T = (c_{i1}, \dots, c_{im_i})$, $i = 1, \dots, n$, is a $1 \times m_i$ vector of random variables.
- (d) $b_i^T = (b_{i1}, \dots, b_{im_i})$, $i = 1, \dots, n$, is a $1 \times m_i$ vector of random variables.
- (e) the joint distribution of the b_{ik} and c_{if} , $k = 1, \dots, m_i, f = 1, \dots, n_{f+1}$, $j = 1, \dots, n$, is known, i.e., the multivariate distribution function of the b_{ik} and c_{if} is assumed to be a known function.

In addition to these five assumptions on the properties of the matrices involved in the problem, some conditions on the way in which the problem is to be solved will be imposed. The nature of these additional assumptions is dictated by the interpretation of the problem in this paper. Problem 10 will be treated as an n -period, or n -stage, problem in which λ_j , the vector of decision rules for the j th stage, is selected with knowledge of all decisions $\lambda_1, \dots, \lambda_{j-1}$ and observations of the vectors of random variables $b_i, c_i, i = 1, \dots, j-1$, but before b_j, c_j , and all random variables of periods succeeding the j th are observed. Thus λ_j will be known exactly (i.e., will be a known number) at the start of the j th period.

In other words, we must select λ_1 , our vector of first-period decisions, before observing the value of the first-period random vector b_1 and the first-period cost vector c_1 . Then, having selected λ_1 and observed b_1 and c_1 , the second-period decision rules λ_2 must be chosen before observing the values of b_2 and c_2 . This process continues for n periods with λ_j depending only on $\lambda_1, b_i, c_i, i = 1, \dots, j-1$, but not on b_j, c_j or $\lambda_1, b_i, c_i, i = j+1, \dots, n$. This interpretation means that the information at any stage (aside from a knowledge of the joint distribution of $b_i, c_i, i = 1, \dots, n$) is limited to a knowledge of the decisions and observations of the preceding stages. Thus it is assumed that—

- (f) the $n_j \times 1$ vector $\lambda_j, j = 1, \dots, n$, is a function of λ_1, b_i, c_i , for $i = 1, \dots, j-1$, but is not a function of the remaining decisions $\lambda_i, i = j+1, \dots, n$, or of random vectors $b_i, c_i, i = j, \dots, n$.

λ_j will, of course, depend implicitly on the decisions and observations of succeeding stages because of the coupling nature of the constraints of problem 10. Assumption f means that λ_j is not an explicit function of b_j, c_j , or $\lambda_1, b_i, c_i, i = j+1, \dots, n$, so that λ_j will be known exactly (i.e., will be a known number) when the j th period decision is made.

As a consequence of this assumption it can be seen that the choice of λ_j affects the objective function of problem 10 not only through the $c_j^T \lambda_j$ term but also through the effect of λ_j on λ_{j+2}, \dots, n . λ_n on the other hand enters the functional only in the $c_n^T \lambda_n$ term.

It will now be shown how assumptions e and f enable the constraints of problem 10 to be written in a more convenient form. The P operator in the

constraints means that the probability is computed using the joint distribution of $b_i, c_i, i, j=1, \dots, n$. Thus, by using a Stieltjes integral, the following can be written:

$$P(b_j \geq A_{j1}X_1 + \dots + A_{jn}X_n) = \int_A dF(\cdot)$$

where $F(\cdot)$ is the joint distribution function of the $b_i, c_i, i, j=1, \dots, n$, and A is the set where $b_j > A_{j1}X_1 + \dots + A_{jn}X_n$.

It is important that it is clearly understood what is meant by the above identity. This is most easily seen by writing out the j 'th of the m equations (as b_j is $m \times 1$) in detail:

$$\begin{aligned} P(b_{j1} &\geq a_{j1}^{(l)}X_1 + \dots + a_{jn}^{(l)}X_n) \\ &= \int_{a^{(l)}}^{\infty} \int dF(b_{11}, \dots, b_{1m_1}, \dots, b_{n1}, \dots, b_{nm_n}, c_{11}, \dots, c_{1n_1}, \dots, c_{n1}, \dots, c_{nn_n}) \end{aligned}$$

where $a_{ji}^{(l)}$ is the i 'th row of A_{jl} and $a^{(l)}$ is the region in the sample space of the random variables involved in the problem over which $b_{j1} > a_{j1}^{(l)}X_1 + \dots + a_{jn}^{(l)}X_n$. Hence the compact notation

$$P(b_j \geq A_{j1}X_1 + \dots + A_{jn}X_n) = \int_A dF(\cdot)$$

represents m identities of the type shown.

By assumption e, $F(\cdot)$ is a known function. Because of assumption f, the set A depends only on $b_i, c_i, i=1, \dots, j-1$, and b_j and not on the remaining b_i and c_i . Consequently the integration in the above identities can be performed by integrating first with respect to b_{j+1}, \dots, b_n , and c_1, \dots, c_n , and then integrating with respect to b_1, \dots, b_j , and c_1, \dots, c_{j-1} . But when integrated with respect to $b_i, c_i, i=j+1, \dots, n$, and c_j , the region of integration is over the entire range of possible values of these vectors of random variables (i.e., the range of all values that the random variables can assume with non-zero probability) and hence the resulting value of the integral is 1. Thus $\int_A dF(\cdot) = \int_A dF_j(\cdot)$, where $F_j(\cdot)$ is the joint distribution function of b_1, \dots, b_j and c_1, \dots, c_{j-1} . Moreover $F_j(\cdot)$ is a known function because $F(\cdot)$ is known.

Through the use of conditional probabilities $\int_A dF_j(\cdot)$ can be written in terms of $\bar{F}_j(\cdot)$, the conditional distribution function of b_j given $b_i, c_i, i=1, \dots, j-1$, and $G_j(\cdot)$, the joint distribution function of $b_i, c_i, i=1, \dots, j-1$. We have, by definition, $\int_A dF_j(\cdot) = \int_{-\infty}^{\infty} (\int_A d\bar{F}_j(\cdot)) dG_j(\cdot)$. Again $\bar{F}_j(\cdot)$ and $G_j(\cdot)$ are known because they can be computed from a knowledge of $F(\cdot)$.

Now $\bar{\sigma}_j, j=1, \dots, n$ is defined as an $m \times 1$ vector of probabilities such that

$$\int_{-\infty}^{\infty} \left(\int_A d\bar{F}_j(\cdot) \right) dG_j(\cdot) \geq a_j \text{ if and only if } \int_A d\bar{F}_j(\cdot) \geq \bar{\sigma}_j$$

The $\bar{\sigma}_j$ will be conditional probabilities depending on the given (i.e., conditional) values of $b_i, c_i, i=1, \dots, j-1$, which are used in determining $\bar{F}_j(\cdot)$. In other words, $\bar{\sigma}_j$ is such that

$$P(b_j \geq A_{j1}X_1 + \dots + A_{jn}X_n) \geq \bar{\sigma}_j \quad (11)$$

if and only if

$$\bar{P}(b_j \geq A_{j1}x_1 + \dots + A_{jj}x_j) \geq \bar{\alpha}_j$$

where the \bar{P} operator means that the probability is computed using the conditional distribution of b_j given $b_i, c_i, i=1, \dots, j-1$, i.e., using $\bar{F}_j(\cdot)$. Thus, for example, if b_j is independent of $b_i, c_i, i=1, \dots, j-1$, the \bar{P} represents probability using the marginal distribution of b_j .

It is clear from the above development that if, instead of having P and σ_j , as in the constraints of problem 10, a similar problem is given whose constraints involve \bar{P} and some set of conditional probabilities $\bar{\alpha}_j, j=1, \dots, n$, the corresponding total probabilities P , along with their associated σ_j , can be obtained by integrating P and $\bar{\alpha}_j$ with respect to $G_j(\cdot)$.

It is important to note in this development that $\bar{\alpha}_j = \sigma_j$ because $F_j(\cdot)$ is the marginal distribution function of b_j and so is $\bar{F}_j(\cdot)$. In other words, $\bar{\alpha}_j$ is a vector of constants and its components do not depend on the observed values of any of the random variables involved in the problem. This agrees with the interpretation of the problem in this paper because x_i must be selected before any of the random variables are observed. Consequently the first-period constraints must have a completely deterministic equivalent. This is precisely the type of interpretation that led to Ben-Israel's theorem¹⁷ and its associated results (see Section 4). In brief, x_i is a zero-order decision rule.

The question of how to determine $\bar{\alpha}_j$ for given σ_j and distribution functions $\bar{F}_j(\cdot)$ and $G_j(\cdot)$ is extremely difficult. Here, however, determining analytical properties of x_j in terms of $\bar{\alpha}_j$ is of chief concern; hence the problem of determining $\bar{\alpha}_j$ will be left for a subsequent paper. However, it must be emphasized that $\bar{\alpha}_j$ can (and in general will) depend on the given values of $b_i, c_i, i=1, \dots, j-1$. This is of particular importance because it greatly enlarges the types of constraints that are permissible in such a model. Elsewhere in the literature it is always assumed that the probabilities with which the constraints must hold are given constants no matter what values of the random variables are observed. Thus a more general problem is being treated than any previously considered in the literature.

Relation 11 and assumption i can be used to prove the following:

Lemma 1. The constraint $P(b_j \geq A_{j1}x_1 + \dots + A_{jj}x_j) \geq \sigma_j$ in problem 10 can be replaced by the equivalent constraint

$$A_{j1}x_1 + A_{j2}x_2 + \dots + A_{jj}x_j \leq \bar{F}_j^{-1}(1-\bar{\alpha}_j)$$

where $\bar{F}_j^{-1}(1-\bar{\alpha}_j)$ is an $m_j \times 1$ vector of the $1-\bar{\alpha}_j$ percentile (or fractile) points of the conditional distribution of b_j given $b_i, c_i, 1 \leq i < j$. The i th component of $\bar{F}_j^{-1}(1-\bar{\alpha}_j)$, $\bar{F}_j^{-1}(1-\bar{\alpha}_{ji})$, is defined by $\bar{F}_j^{-1}(1-\bar{\alpha}_{ji}) = \text{maximum } \{y : \bar{F}_j(y) \leq 1-\bar{\alpha}_{ji}\}$.

The use of a maximum in the definition of $\bar{F}_j^{-1}(1-\bar{\alpha}_{ji})$ is justified because the assumption is that $\bar{F}_j(\cdot)$ is left continuous.[†] The reason the vector inequality $\bar{F}_j(Y) \leq 1-\bar{\alpha}_j$ does not necessarily define Y uniquely is due to the fact that some of the random variables involved in $F_j(\cdot)$ may be discrete and that some of the components of $1-\bar{\alpha}_j$ may be zero. In general each component, y_i ,

[†]For a further discussion of left continuous distribution functions see Gnedenko, (Ref 24, Ch 4).

of Y can take on any value in an interval. If this interval is denoted by $y_i(1-\bar{\alpha}_{ji})$ and $\hat{y}_i(1-\bar{\alpha}_{ji})$, $\hat{y}_i(1-\bar{\alpha}_{ji})$ respectively represent the smallest and largest numbers in the interval $y_i(1-\bar{\alpha}_{ji})$, then the i th component of $\bar{F}_j^{-1}(1-\bar{\alpha}_j)$ is given by $\hat{y}_i(1-\bar{\alpha}_{ji})$.

It is assumed throughout this discussion that $\bar{F}_j^{-1}(1-\bar{\alpha}_{ji})$ is well defined for all i, j . Thus, for example, if $\bar{F}_j(\cdot)$ is the distribution function of a normal random variable then it is implicitly assumed that no component of $\bar{\alpha}_j$ is 1, so that $\bar{F}_j^{-1}(1-\bar{\alpha}_{ji})$ does not take on the value of $-\infty$ for any i . Similarly it can be seen that $\bar{\alpha}_{ji}$ cannot be zero for any i or j , since in this case $1-\bar{\alpha}_{ji} = 1$ and so $\hat{y}_i(1-\bar{\alpha}_{ji}) = +\infty$, no matter what distribution $\bar{F}_j(\cdot)$ is used. This is, however, no restriction on our problem, since $\bar{\alpha}_{ji} = 0$ means that the corresponding constraint can be ignored because it will be satisfied for any choice of decision rules X_1, \dots, X_j .

Proof. Assume that $X_1, 1 \leq i \leq j$ are decision rules that satisfy $P(b_j \geq A_{j1}X_1 + \dots + A_{jj}X_j) \geq \alpha_j$. Then

$$\bar{P}(b_j \geq A_{j1}X_1 + \dots + A_{jj}X_j) \geq \bar{\alpha}_j \text{ (by relation 11),}$$

or

$$1 - \bar{P}(b_j < A_{j1}X_1 + \dots + A_{jj}X_j) \geq \bar{\alpha}_j,$$

or

$$\bar{P}(b_j < A_{j1}X_1 + \dots + A_{jj}X_j) \leq 1 - \bar{\alpha}_j.$$

But $\bar{P}(b_j < A_{j1}X_1 + \dots + A_{jj}X_j) = \int_B d\bar{F}_j(\cdot)$ where B is the set for which $b_j < A_{j1}X_1 + \dots + A_{jj}X_j \in Z$. Also, by assumptions a and f, $A_{j1}X_1 + \dots + A_{jj}X_j$ is not a function of b_j , hence Z does not depend on the still-to-be-observed value of b_j . Consequently we have

$$\int_B d\bar{F}_j(\cdot) = \int_{-\infty}^Z d\bar{F}_j(\cdot) = \bar{F}_j(Z) = \bar{F}_j(A_{j1}X_1 + \dots + A_{jj}X_j).$$

Thus the j th constraint of problem 10 requires that

$$\bar{F}_j(A_{j1}X_1 + \dots + A_{jj}X_j) \leq 1 - \bar{\alpha}_j.$$

But any distribution function is a nondecreasing function of its argument, so that $\bar{F}_j(A_{j1}X_1 + \dots + A_{jj}X_j) \leq 1 - \bar{\alpha}_j$ implies

$$A_{j1}X_1 + \dots + A_{jj}X_j \leq \bar{F}_j^{-1}(1 - \bar{\alpha}_j). \quad (12)$$

Since all the steps of the proof are completely reversible, relation 12 implies

$$P(b_j \geq A_{j1}X_1 + \dots + A_{jj}X_j) \geq \alpha_j.$$

Thus the lemma is proved.[†]

Using Lemma 1, problem 10 is written in the following equivalent form:

[†]Note the similarity between Lemma 1 and Ben-Israel's theorem.¹⁷

maximize

$$E(c_1^T X_1 + c_2^T X_2 + c_3^T X_3 + \dots + c_n^T X_n)$$

subject to

$$\begin{aligned} A_{11} X_1 &\leq \bar{F}_1^{-1}(1-\bar{\alpha}_1), \\ A_{21} X_1 + A_{22} X_2 &\leq \bar{F}_2^{-1}(1-\bar{\alpha}_2), \\ A_{31} X_1 + A_{32} X_2 + A_{33} X_3 &\leq \bar{F}_3^{-1}(1-\bar{\alpha}_3), \\ &\vdots \\ A_{n1} X_1 + A_{n2} X_2 + A_{n3} X_3 + \dots + A_{nn} X_n &\leq \bar{F}_n^{-1}(1-\bar{\alpha}_n). \\ X_j &\geq 0 \quad j=1, \dots, n. \end{aligned} \tag{13}$$

The E operator in the functional of problems 13 and 10 means that the expected value is being taken with respect to the joint distribution of $b_1, c_1, \dots, b_n, c_n$, $j=1, \dots, n$. By using the linearity of the objective function in $X_j, j=1, \dots, n$, the following may be written:

$$E\left(\sum_{j=1}^n c_j^T X_j\right) = \sum_{j=1}^n \int_{-\infty}^{\infty} c_j^T X_j dF(\cdot).$$

However, this expected-value operation can be simplified by making use of assumption f. Since X_j is a function of $b_1, c_1, \dots, b_{j-1}, c_{j-1}, \dots, b_j$, only,

$$\begin{aligned} E(c_j^T X_j) &= \int_{-\infty}^{\infty} c_j^T X_j dF(\cdot) = \int_{-\infty}^{\infty} c_j^T X_j \left[\int_{-\infty}^{\infty} d\hat{F}_j(\cdot) \right] d\hat{F}_j(\cdot) \\ &= \int_{-\infty}^{\infty} c_j^T X_j d\hat{F}_j(\cdot) \end{aligned}$$

where $d\hat{F}_j(\cdot)$ is the conditional distribution function of $b_1, \dots, b_{j-1}, c_1, \dots, c_{j-1}$, given $b_1, \dots, b_{j-1}, c_1, \dots, c_j$, and $\hat{F}_j(\cdot)$ is the joint distribution function of $b_1, \dots, b_{j-1}, c_1, \dots, c_j$. Thus, in computing $E(c_j^T X_j)$, $c_j^T X_j$ is integrated by using only $F_j(\cdot)$ rather than $F(\cdot)$.

7. MATRIX THEORY

In the following discussion of the optimal decision rules use is made of a generalized inverse of a matrix. The properties required of such inverses are listed below. Proofs may be found in Charnes and Kirby²⁵ and Rao.²⁶

Definition. Let A be any $m \times n$ matrix. Then any $n \times m$ matrix A^* such that $AA^*A = A$ is called a generalized inverse of A .

It can be shown that such an inverse always exists and is unique if and only if $m = n$.

Lemma 2. Let A be an $m \times n$ matrix. Then there exists an A^* such that $AA^* = I$ if and only if $\text{rank}(A) = m$, i.e., A has a right inverse if and only if it has full row rank.

Lemma 3. The matrix equation $AX = b$ is consistent if and only if $AA^*b = b$.

Lemma 4. Let the matrix equation $AX = b$ be consistent. Then the general solution is given by $X = A^*b + (I - A^*A)Y$ where Y is an arbitrary $n \times 1$ vector and I is the $n \times n$ identity matrix.

Using these lemmas, a necessary and sufficient condition can be given for the existence of a vector X satisfying the set of inequalities $AX \leq b$, $X \geq 0$. For, there exists $X \geq 0$ such that $AX \leq b$.

- there exists $X, S \geq 0$ such that $AX + S = b$ where S is an $n \times 1$ vector
- there exists $X, S \geq 0$ such that $AX = b - S$, $AA^*(b - S) = b - S$ by using

Lemma 3.

- there exists an arbitrary vector Y and a vector $S \geq 0$ such that

$$AA^*(b - S) = b - S$$

$$A^*(b - S) \geq (A^*A - I) Y$$

by using **Lemma 4**. In this case $X = A^*(b - S) + (I - A^*A) Y$. Therefore there exists a finite $n \times 1$ vector X satisfying the set of inequalities $AX \leq b$, $X \geq 0$ if and only if there exists a finite $n \times 1$ vector Y and $n \times 1$ vector S , $S \geq 0$ such that $AA^*(b - S) = b - S$ and $A^*(b - S) \geq (A^*A - I) Y$. From **Lemma 2** it can be seen that if A has full row rank, the condition $AA^*(b - S) = b - S$ becomes an identity and it is only required that there exist vectors Y and $S \geq 0$ such that $A^*(b - S) \geq (A^*A - I) Y$.

8. REMARKS ABOUT THE CONSISTENCY OF THE CONSTRAINTS

A significant property of chance-constrained programming models is that the constraints generally represent the "intentions" of management rather than hard and fast "rules." In other words, the constraints represent "bounds" inside which management would like to operate "most of the time" rather than "all the time." The decision rules resulting from solving a chance-constrained problem are designed to give guidelines rather than definite plans of action for management operation and decision, subject to qualifications surrounding the controls involved in implementing the rules. Thus the optimal decision rules for our problem may impute an action that, because of exceptional circumstances, cannot actually be taken.

Another major feature of the chance-constrained programming concept is that, in contrast to so-called "linear programming under uncertainty," the constraints (e.g., $\sum_{j=1}^n a_{ij} x_j \leq b_i$, as in relation 3) need not hold with probability 1. This generalization is important on several realistic grounds. First, it may not be possible, or even desirable, to specify actions in every conceivable circumstance, i.e., for every possible value of the random variables. For example, in the heating-oil problem (see Section 2), certain combinations of events (which really did occur once or twice) were impossible for the company, acting alone, to handle either physically or economically. Second, constraints that have the nature of "policies" are not really intended to apply in every instance but only "almost all the time." Hence the constraints of the problem in Section 2 say that the supply of heating oil must exceed the customer demand, not all the time but 100 α_1 percent of the time, where α_1 is some number close to but not equal to 1. Third, by their nature, policies (as opposed to definite rules) need not spell out in advance the actions that will be taken in exceptional circumstances.

Chance-constrained programming allows for such exceptional circumstances in a very easy, natural manner, particularly when the admissible classes are defined in some straightforward analytical manner, e.g., as linear decision rules. However, questions arise when one wishes to consider "all possible" or "all possibly desirable" stochastic actions in time sequence. It may then be seen, on reflection, that an essential feature of the analytically specified admissible class is that the analytic description carries one through sets of random events that imply constraint violations or the impossibility of actual, as opposed to imputed, further actions in the time sequence. For example, in the heating-oil problem the linear decision rules defined a value of R_j no matter how large the values observed for $S_i, i = 1, \dots, j-1$. Clearly then, if extremely large demands were observed in periods $1, \dots, j-1$, it would be physically impossible for the company to produce enough oil to meet the amount given by R_j . This would represent an exceptional circumstance and the company would not (and could not) follow the decision for period j imputed by the linear decision rule. Thus it is important to consider the question of what analytic specifications should be made to carry one through such situations in solving for optimal rules.

In addition this example shows that if constraints that place an upper bound on $R_j, j = 1, \dots, n$ are added to problem 4 it is possible that the constraints will be inconsistent for some values of $S_i, i = 1, \dots, n$. Therefore discussion of the difficulties involved when the constraints of the problem are inconsistent must precede the establishment of certain properties of the decision rules that are optimal for this problem.

The question of inconsistency is of significance because, as has been illustrated, one of the major (and frequently overlooked and misunderstood) properties of chance-constrained models is that points for which the constraints are inconsistent may well exist in the sample space of the random variables involved in the problem. In other words there may be points $(b_1, \dots, b_n, c_1, \dots, c_n)$ in the sample space of the random variables $b_i, c_i, i = 1, \dots, n$, for which there do not exist decision rules $X_j, j = 1, \dots, n$ that satisfy the constraints of problem 13. Thus it is possible that for some values of the random variables $b_i, c_i, i = 1, \dots, j-1$, and first $j-1$ decisions, $X_i, i = 1, \dots, j-1$, there will not exist an $n-j+1$ vector $X_j > 0$ satisfying $A_{jj} X_j \leq \bar{F}_j^{-1}(1-\bar{\alpha}_j) - A_{j,j-1} X_{j-1} - \dots - A_{1j} X_1$. In such cases the question arises as to the meaning of the expectation operator in the functional. How is $E(c_j^T X_j)$ calculated over those sample points for which no feasible X_j exists?

One method of resolving this difficulty is to replace $\sum_{j=1}^n E(c_j^T X_j)$ in the functional by $\sum_{j=1}^n E_j(c_j^T X_j)$, where E_j means that the expected value is taken over only those sample points for which there exist decision rules X_1, \dots, X_j , satisfying the first j constraints of problem 13. Thus

$$E(c_j^T X_j) = \int_{-\infty}^{\infty} c_j^T X_j d\tilde{F}_j(\cdot)$$

but

$$E_j(c_j^T X_j) = \int_{\Delta_j} c_j^T X_j d\tilde{F}_j(\cdot)$$

where Δ_j denotes the region where the first j constraints of problem 13 are consistent.

If $\bar{\Delta}_j$ denotes the complement of Δ_j , it can be seen that $E(c_j^T X_j)$ and $E_j(c_j^T X_j)$ are related by

$$E_j(c_j^T X_j) = E(c_j^T X_j) - \int_{\bar{\Delta}_j} c_j^T X_j d\tilde{F}_j(\cdot).$$

By applying the mean value theorem we obtain the equation

$$E(c_j^T X_j) - E_j(c_j^T X_j) = k P(\bar{\Delta}_j)$$

where k is a constant and

$$P(\bar{\Delta}_j) = \int_{\bar{\Delta}_j} d\tilde{F}_j(\cdot).$$

Thus we see that whenever $P(\bar{\Delta}_j)$ is small, the use of E_j provides a close approximation to E .

The concept of an analytic decision-rule class under chance constraints originated by Charnes and Cooper^{2,10,27,28} provides a direct and simple means of specifying the decision rules across the inconsistent points of the problem. In this case the analytic description of the class permits X_j to be defined throughout all points of the sample space of the random variables (see the preceding discussion and the example of Section 2). Therefore, when a class of decision rules is specified, X_j becomes defined throughout $\bar{\Delta}_j$ and the use of the E operator again becomes meaningful.

One way of defining X_j in $\bar{\Delta}_j$ is to specify $X_j = 0$ in this region. The effect of this definition is to make $E_j(c_j^T X_j) = E(c_j^T X_j)$. Moreover use can be made of the material developed in Section 7 to specify, a priori, the region Δ_j over which $X_j = 0$ will be put. It is known from Section 7 that there exists a decision rule $X_j > 0$ satisfying $A_{jj} X_j \leq \bar{F}_j^{-1}(1-\bar{\alpha}_j) = A_{jj-1} X_{j-1} + \dots + A_{j1} X_1 + S_j$ if and only if there exist vectors Y_j and $S_j > 0$ satisfying

$$\begin{aligned} A_{jj} A_{jj}^{-\#} (\bar{F}_j^{-1}(1-\bar{\alpha}_j) - A_{jj-1} X_{j-1} - \dots - A_{j1} X_1 - S_j) \\ = \bar{F}_j^{-1}(1-\bar{\alpha}_j) - A_{jj-1} X_{j-1} - \dots - A_{j1} X_1 - S_j. \end{aligned}$$

and

$$A_{jj}^{-\#} (\bar{F}_j^{-1}(1-\bar{\alpha}_j) - A_{jj-1} X_{j-1} - \dots - A_{j1} X_1 - S_j) \geq (A_{jj}^{-\#} A_{jj} - I) Y_j$$

and in this region

$$X_j = A_{jj}^{-\#} (\bar{F}_j^{-1}(1-\bar{\alpha}_j) - A_{jj-1} X_{j-1} - \dots - A_{j1} X_1 - S_j) + (I - A_{jj}^{-\#} A_{jj}) Y_j.$$

If D_j denotes the region where a feasible X_j exists, it can be seen that Δ_j , the set of points for which the first j constraint's are consistent, is given by $\bigcap_{i=1}^j D_i$. Hence $\bar{\Delta}_j$, the complement of Δ_j , equals $\bigcup_{i=1}^j \bar{D}_i$, where \bar{D}_i is the complement of D_i .

The specification of X_j as constant ($= 0$) over its region of inconsistency makes it possible to extend the results obtained below. In Theorem 2 it will be proved that in the region where the constraints are consistent the optimal decision rule for X_j gives X_j as a piecewise linear function of $\bar{F}_j^{-1}(1-\bar{\alpha}_j)$, and

x_{j-1}, \dots, x_1 . Therefore specifying $x_j = 0$ in $\bar{\Delta}_j$ means that the optimal decision rules become piecewise linear everywhere.

Another method of resolving the difficulty of inconsistency that could yield values of $x_j \neq 0$ in $\bar{\Delta}_j$ is to modify the formulation of the constraints of the problem so that the right-hand side of the constraints in problem 13 is operated on by a projection into the range of the A_{jj} operator. This projection would be the identity operator when consistency holds and would yield some near point in the range of A_{jj} when the j th constraint is inconsistent (if the Penrose-Moore inverse²⁹ were used in the projection operator, the near point would be the nearest feasible point in the Euclidean sense).† With this change in the constraints, the E operator could then be applied without modification and the resulting functional value of $E(c_j^T x_j)$ would differ from $E_j(c_j^T x_j)$ by, at most, something on the order of $P(\bar{\Delta}_j)$.

Regardless of the procedure used to resolve the difficulty of inconsistent sample points, the problem of determining the optimal decision rule for x_j is reduced to finding the optimal rule when the constraints are consistent. At those sample points where sufficient consistency exists for the constraints and the functional to be meaningful, the argument of the following section is valid.

9. OPTIMAL CLASS OF DECISION RULES

Let $(b_1, \dots, b_n, c_1, \dots, c_n)$ be any point in the sample space of the vectors of random variables $b_i, c_{j,i}$, $j = 1, \dots, n$, for which there is consistency, i.e., let this sample point be in $\cap_{i=1}^n D_i$. Suppose that for each such consistent point the decision rules are found to be x_1^*, \dots, x_n^* , maximizing $\sum_{j=1}^n c_j^T x_j$ subject to the constraints of problem 13. Then x_j^* , $j = 1, \dots, n$, are the optimal decision rules for the problem. This follows from the fact that for any other feasible decision rules, x_j , $j = 1, \dots, n$, for problem 13, $\sum_{j=1}^n c_j^T x_j > \sum_{j=1}^n c_j^T x_j^*$ results from the definition of x_j^* , $j = 1, \dots, n$. This implies that $E(\sum_{j=1}^n c_j^T x_j^*) > E(\sum_{j=1}^n c_j^T x_j)$, thus establishing the optimality of x_j^* , $j = 1, \dots, n$, for points where consistency is present.

Theorem 2

The optimal decision rule for x_j in problem 13 is of the form

$$x_j = \text{a piecewise linear function of the } \bar{F}_j^{-1}(1 - \bar{a}_j) \text{ and } x_{j-1}, \dots, x_1.$$

Proof. By assumption f_j , x_j , $j = 1, \dots, n$, influences the objective function through the terms $c_j^T x_j$, $c_{j+1}^T x_{j+1}$, \dots , $c_n^T x_n$. Hence the effect of the choice of x_n appears in the functional only in the $c_n^T x_n$ term. Consequently the first step is to find the optimal decision rule for x_n as a function of x_{n-1}, \dots, x_1 and then proceed recursively by finding the optimal x_{n-1} as a function of x_{n-2}, \dots, x_1 until the problem is ultimately reduced to finding the optimal x_1 .

† See Ben-Israel and Charnes³⁰ and other references on Penrose-Moore inverses given in Section 2.

To avoid the detailed computation for each stage the theorem will be proved by induction on ℓ where $\ell = n + 1 - \ell$. After proving the theorem for $\ell = 1$, i.e., after proving that the optimal X_n is given as a piecewise linear function of $\bar{F}_n^{-1}(1-\bar{\alpha}_n)$, X_{n-1}, \dots, X_1 , it is then assumed for induction that the theorem is true for $\ell = k$, i.e., for X_{n-k+1}, \dots, X_n . It is then proved true for $\ell = k + 1$, i.e., for X_{n-k} .

Let $(b_1, \dots, b_n, c_1, \dots, c_n)$ be any consistent point in the space of possible values for the vectors of random variables, $b_i, c_j, i, j = 1, \dots, n$. For this sample point the problem of maximizing $\sum_{j=1}^n c_j^T X_j$ subject to the constraints of problem 13 will be solved and thus the optimal decision rules $X_j^*, j = 1, \dots, n$, will be obtained.

Let X_1, \dots, X_{n-1} be specific values for the first $n-1$ decision rules. Then our problem becomes one of determining a decision rule for X_n that solves

maximize

$$c_n^T X_n$$

subject to

$$A_{nn} X_n \leq \bar{F}_n^{-1}(1-\bar{\alpha}_n) = A_{n,n-1} X_{n-1} + \dots + A_{n,1} X_1. \quad (14)$$

$$X_n \geq 0.$$

But because $X_j, j = 1, \dots, n-1$ is specified and the focus is on a single point in the sample space, the right-hand side of the constraints in problem 14 is a known vector. Moreover by assumption all A_{ij} are constant matrices; hence problem 14 is a linear programming problem and so has a dual. Its dual is

minimize

$$\omega_n^T [\bar{F}_n^{-1}(1-\bar{\alpha}_n) - A_{n,n-1} X_{n-1} - \dots - A_{n,1} X_1]$$

subject to

$$\begin{aligned} \omega_n^T A_{nn} &\leq c_n^T \\ \omega_n &\geq 0. \end{aligned} \quad (15)$$

If problem 14 has a finite optimal solution X_n^* , then the convex set of feasible solutions to problem 15 has a finite number of extreme points, denoted by $\omega_{ni}^T, i = 1, \dots, N_n$; and, from the dual theorem of linear programming, it is concluded that

$$c_n^T X_n^* = \min_{i=1, \dots, N_n} \left\{ \omega_{ni}^T [\bar{F}_n^{-1}(1-\bar{\alpha}_n) - A_{n,n-1} X_{n-1} - \dots - A_{n,1} X_1] \right\} \quad (16)$$

Since there is only a finite number N_n of extreme points, problem 16 implies that X_n^* is a piecewise linear function of $\bar{F}_n^{-1}(1-\bar{\alpha}_n) - A_{n,n-1} X_{n-1} - \dots - A_{n,1} X_1$ with, at most, N_n pieces. Hence X_n^* is piecewise linear in $\bar{F}_n^{-1}(1-\bar{\alpha}_n)$ and X_{n-1}, \dots, X_1 . Thus the theorem is true for $\ell = 1$.

Now assume for induction that the theorem is true for $\ell = k$. Then there exist functions $H_j(\bar{F}_n^{-1}(1-\bar{\alpha}_j), X_{j-1}, \dots, X_1), j = n-k+1, \dots, n$, which are piecewise linear in $\bar{F}_n^{-1}(1-\bar{\alpha}_j), X_{j-1}, \dots, X_1$, and are such that $X_j^* = H_j, j = n-k+1, \dots, n$. Thus problem 13 can be written as

maximize

$$E(c_1^T \lambda_1 + \dots + c_{n-k}^T \lambda_{n-k} + c_{n-k+1}^T H_{n-k+1} + \dots + c_n^T H_n)$$

subject to

$$\begin{aligned} A_{11}\lambda_1 &\leq \bar{F}_1^{-1}(1-\bar{\alpha}_1), \\ A_{21}\lambda_1 + A_{22}\lambda_2 &\leq \bar{F}_2^{-1}(1-\bar{\alpha}_2), \\ &\vdots \\ A_{n-k,1}\lambda_1 + A_{n-k,2}\lambda_2 + \dots + A_{n-k,n-k}\lambda_{n-k} &\leq \bar{F}_{n-k}^{-1}(1-\bar{\alpha}_{n-k}), \\ \lambda_j &\geq 0, \quad j=1, \dots, n-k. \end{aligned} \tag{17}$$

Again let $(b_1, \dots, b_n, c_1, \dots, c_n)$ be a particular sample point and let $\lambda_1, \dots, \lambda_{n-k-1}$ be specified values for the first $n-k-1$ decision rules. Then λ_{n-k}^* is obtained by solving the following:

maximize

$$c_{n-k}^T \lambda_{n-k} + \sum_{j=n-k+1}^n c_j^T H_j$$

subject to

$$\begin{aligned} A_{n-k,n-k}\lambda_{n-k} &\leq \bar{F}_{n-k}^{-1}(1-\bar{\alpha}_{n-k}) = A_{n-k,n-k-1}\lambda_{n-k-1} + \dots + A_{n-k,1}\lambda_1, \\ \lambda_{n-k} &\geq 0. \end{aligned} \tag{18}$$

The functional in problem 18 is piecewise linear in λ_{n-k} because of the piecewise linearity of the $H_j, j=n-k+1, \dots, n$. Moreover the optimal λ_{n-k} can be obtained by first solving problem 18 over each piece where the functional is linear in λ_{n-k} and then selecting the piece that is optimal. That is, the solution space of problem 18 can be divided into sets K_1, \dots, K_N , such that in any set K_r the functional in problem 18 is linear in λ_{n-k} . Problem 18 is then solved for λ_{n-k}^* , the optimal λ_{n-k} in $K_r, r=1, \dots, N$, and finally λ_{n-k}^* is selected such that

$$c_{n-k}^T \lambda_{n-k}^* + \sum_{j=n-k+1}^n c_j^T H_j(\lambda_{n-k}^*) = \max \left\{ c_{n-k}^T \lambda_{n-k,r}^* + \sum_{j=n-k+1}^n c_j^T H_j(\lambda_{n-k,r}^*) \right\}, \quad r=1, \dots, N.$$

Since for any set K_r the functional in problem 18 is linear in λ_{n-k} , the problem in K_r is

maximize

$$\Sigma_{n-k}^T \lambda_{n-k}$$

subject to

$$\begin{aligned} A_{n-k,n-k}\lambda_{n-k} &\leq \bar{F}_{n-k}^{-1}(1-\bar{\alpha}_{n-k}) = A_{n-k,n-k-1}\lambda_{n-k-1} + \dots + A_{n-k,1}\lambda_1, \\ \lambda_{n-k} &\geq 0 \end{aligned} \tag{19}$$

where Σ_{n-k}^T is a $1 \times n_{n-k}$ matrix whose components depend on the sample point $(b_1, \dots, b_n, c_1, \dots, c_n)$ and the given values of $\lambda_1, \dots, \lambda_{n-k-1}$.

Because problem 19 is a linear programming problem like problem 14, by writing its dual and proceeding exactly as was done in going from problem 14 to problem 16, it can be seen that λ_{n-k}^* is a piecewise linear function of

$\bar{F}_{n-k}^{-1}(1-\bar{\alpha}_{n-k})$, X_{n-k-1}, \dots, X_1 . Since this is true for each region K_i over which the functional in problem 18 is linear, it is concluded that X_{n-k}^* is indeed piecewise linear in $\bar{F}_{n-k}^{-1}(1-\bar{\alpha}_{n-k})$, X_{n-k-1}, \dots, X_1 .

Thus, by induction, the theorem is proved.

Corollary 1. The optimal decision rule for X_j , $j = 1, \dots, n$, in problem 13 is of the form $X_j =$ a piecewise linear function of the $\bar{F}_j^{-1}(1-\bar{\alpha}_j)$, $j = 1, \dots, n$.

10. OPTIMAL DECISION RULES FOR SOME SPECIAL CASES

The results of the preceding section apply to any model satisfying assumptions a through f. These assumptions are, however, very general and so it is instructive to look in closer detail at certain special cases that yield particularly interesting results. More specifically the chief concern will be the case in which problem 10 is an n -stage linear-programming-under-uncertainty problem (i.e., $a_{ij} = 1$, $i = 1, \dots, m_i$, $j = 1, \dots, n$).

However, another important special case of problem 10 is considered in Ref 5. This is the case in which each A_{ij} matrix, $i > j$, $i, j = 1, \dots, n$, is a 1×1 matrix. In other words, each period generates exactly one new constraint (instead of m_i new constraints as in problem 10). Charnes and Kirby⁵ have shown that there is much more to say about the solution of such problems than has been said in Theorem 2. In particular an explicit method of obtaining the optimal decision rules under certain circumstances is shown.

To begin with, however, it is observed that Corollary 1 also implies the following.

Corollary 2. The optimal decision rule for X_j , $j = 1, \dots, n$, in problem 13 is piecewise linear in the b_i , c_i , $i = 1, \dots, j-1$, if and only if $\bar{F}_j^{-1}(1-\bar{\alpha}_j)$, $j = 1, \dots, l$, is piecewise linear in b_i , c_i , $i = 1, \dots, l-1$.

As an example of a situation in which Corollary 2 is applicable, consider the case where b_1, \dots, b_n , c_1, \dots, c_n are jointly normally distributed, and $\bar{\alpha}_{ij}$ is a constant for all j , i , and so does not depend on the given values of b_i , c_i , $i = 1, \dots, j-1$. Then $\bar{F}_j(\cdot)$ is the distribution function of a normal random variable with mean m and variance δ^2 . Moreover m is a linear function of the random variables that are given in the conditional distribution. Thus m is a linear function of b_i , c_i , $i = 1, \dots, j-1$. Also δ^2 is a constant. Therefore $\bar{F}_j(y_j) = \Phi(y_j - m / \delta)$ can be written where $\Phi(\cdot)$ is the distribution function of a normal random variable with mean 0 and variance 1. Thus

$$\bar{F}_j(y_j) = 1 - \bar{\alpha}_{jj} = y_j = \bar{F}_j^{-1}(1 - \bar{\alpha}_{jj}) = \delta \Phi^{-1}(1 - \bar{\alpha}_{jj}) + m.$$

But $\bar{\alpha}_{jj}$ is, by hypothesis, a constant. Hence $\Phi^{-1}(1 - \bar{\alpha}_{jj})$ is a constant, and the linearity of m implies the linearity of $\bar{F}_j^{-1}(1 - \bar{\alpha}_{jj})$. Therefore each component of $\bar{F}_j^{-1}(1 - \bar{\alpha}_{jj})$, $j = 1, \dots, n$, is a linear function of b_i , c_i , $i = 1, \dots, j-1$.

To give examples where the fractile points are linear in the given sample points and the $\bar{\alpha}_j$ are not constant is fairly difficult. In fact it is clear that X_j^* will not, in general, be piecewise linear in b_i , c_i , $i = 1, \dots, j-1$.

Another special case of this corollary falls into the category of linear programming under uncertainty.[†] Suppose that $a_{ij} = 1$, $i = 1, \dots, n$. Then by

[†]See Charnes, Cooper and Thompson²¹ and G. B. Dantzig et al.^{20, 31-34}

our definition of $\bar{\alpha}_j$, there must also be $\bar{\alpha}_j = 1$, $j = 1, \dots, n$. Assume also that the conditional distribution of b_j , $j = 1, \dots, n$, given b_i, c_i , $i = 1, \dots, j-1$, is such that its 0th fractile point does not depend on the given values of b_i, c_i , $i = 1, \dots, j-1$. Then $\bar{F}_j^{-1}(1-\bar{\alpha}_j) = \bar{F}_j^{-1}(0)$, $j = 1, \dots, n$, becomes a constant vector the components of which are known before any observations have been made. The solutions space of problem 13 is therefore a fixed convex set whose bounding hyperplanes do not depend on the random variables involved in the problem. Thus we are led to the following:

Theorem 3

Assume $\alpha_j = 1$, $j = 1, \dots, n$.

Assume $\bar{F}_j^{-1}(0)$, $j = 1, \dots, n$, is a vector of constants.

Then the optimal decision rule for X_j in problem 13 is such that X_j is piecewise constant.

Proof. By Corollary 1 the optimal X_j is piecewise linear in $\bar{F}_j^{-1}(1-\bar{\alpha}_j)$, $j = 1, \dots, n$. Therefore X_j is piecewise linear in the $\bar{F}_j^{-1}(0)$, $i = 1, \dots, j-1$. But $\bar{F}_j^{-1}(0)$ is a constant by assumption. Hence, X_j is piecewise constant.

This result was also obtained in a different manner by Charnes, Cooper and Thompson²¹ for the n -stage linear-programming-under-uncertainty model (see Theorem 1).

The assumption used above that $\bar{F}_j^{-1}(0)$ be a constant vector is not as restrictive as it might appear at first glance. Indeed this assumption is true for a very large class of distribution functions $\bar{F}_j(\cdot)$. In fact in Theorem 3 only such distributions are eliminated as $\bar{F}_j(\cdot)$, the distribution function of a random variable that is uniform over (a, b) where a depends on b_i, c_i , $i = 1, \dots, j-1$. However, if a is a constant, b can depend on b_i, c_i , $i = 1, \dots, j-1$, and $\bar{F}_j(\cdot)$ will still satisfy the requirement that $\bar{F}_j^{-1}(0)$ be a constant.

This discussion also illustrates one of the major weaknesses of linear programming under uncertainty, i.e., that for many distributions the constraints of problem 13 are inconsistent everywhere. Suppose again that b_i, c_i , $i = 1, \dots, n$, are jointly normally distributed. Then $F_j(\cdot)$ is the distribution function of a normal random variable, hence $\bar{F}_j^{-1}(0) = -\infty$. Consequently there do not exist decision rules X_j , $j = 1, \dots, n$, that satisfy the constraints of problem 13, no matter what sample points are observed.

Further results are obtained on the linear-programming-under-uncertainty problem by restricting the above model still further. To do this it is assumed that c_j , $j = 1, \dots, n$ is a vector of constants rather than random variables and then problem 13 is shown to reduce to an ordinary linear programming problem. When c_j , $j = 1, \dots, n$, are constants, the E operator means that the expected value of $\sum_{j=1}^n c_j^T X_j$ is taken, using the joint distribution of b_1, \dots, b_n . Therefore if problem 13 is solved for a given sample point (b_1, \dots, b_n) , the optimal decision rules X_j^* , $j = 1, \dots, n$ will be obtained for this particular sample point. But $\bar{\alpha}_j = 1$, $j = 1, \dots, n$, and by assumption $\bar{F}_j^{-1}(0)$ is a vector of constants; hence the constraint set of problem 13 does not depend on the sample point (b_1, \dots, b_n) . Moreover since the c_j are constant, problem 13 is now completely deterministic. Thus X_j^* , $j = 1, \dots, n$ will be optimal for all sample points and these optimal decision rules can be found by solving the ordinary linear programming problem.

maximize

$$\sum_{j=1}^n c_j^T x_j$$

subject to

$$A_{11} X_1 \leq \bar{F}_1^{-1}(0).$$

$$A_{21}x_1 + A_{22}x_2 \leq F_2^{-1}(0),$$

$$A_{n+1}X_1 + A_{n+2}X_2 + \dots + A_{nn}X_n \leq \bar{F}_n^{-1}(0),$$

$$V_j \geq 0, \quad j=1, \dots, n.$$

Another example of a situation in which the solutions space of problem 13 does not depend on the observed values of $b_1, c_1, \dots, b_n, c_n$ can be obtained by supposing that $b_j, j = 1, \dots, n$ is stochastically independent of $b_i, c_i, i = 1, \dots, j-1$ and that $\alpha_j, j = 1, \dots, n$ is a constant vector. Then $F_j^{-1}(1-\bar{\alpha}_j)$ is a constant vector, since it represents the $1-\bar{\alpha}_j$ th fractile point of the marginal distribution of b_j . But by assumption e this marginal distribution is known, so $F_j^{-1}(1-\bar{\alpha}_j)$ can be computed before taking any observations. Thus the solutions space of problem 13 does not depend on the observed random variables.

Even when b_j is dependent on the b_1, c_1, \dots, b_{j-1} , $\bar{F}_j^{-1}(1-\bar{\alpha}_j)$ is a vector of constants because it represents the $m_1 + 1$ vector of $1-\bar{\alpha}_j$ percentile points of the marginal distribution of b_j , and $\bar{\alpha}_j$, α_j , is constant.

It is interesting to note that if c_j , $j = 1, \dots, n$ is a constant vector, if b_i , $i = 1, \dots, m$ is independent of the joint distribution of b_i , $i = 1, \dots, j-1$, and if \bar{a}_j , $j = 1, \dots, n$ is constant, then, by an argument analogous to that used above, it can be concluded that problem 13 becomes an ordinary linear programming problem.

11. EXAMPLES

To illustrate the results obtained two examples are presented. In the first example, problem 20, the constraints are consistent for all sample points and all values of $\bar{\alpha}$, in $[0, 1]$. In the second case, problem 23, it can be seen how a well-posed problem can have sample points for which the constraints are inconsistent.

In both these examples the various $\bar{\sigma}_i$ will not be explicitly written as functions of $b_i, r_i, i=1, \dots, l-1$. Thus the problem will be solved treating $\bar{\sigma}_i$ as an arbitrary function of its conditional random variables, subject only to the restriction that $0 < \bar{\sigma}_{ij} < 1$ for all j and i . In addition it will be assumed that our problems are given in the form of problem 13 rather than problem 10, so that the constraints involve \bar{P} and $\bar{\sigma}_i$ and not P and σ_i .

Example 1

As our first example the following two-stage problem in which $x_1 \in x_1$, and $x_2 \in x_2$, are 1 × 1 vectors (i.e., x_1, x_2 are single variables) will be solved:

maximize

$$E(c_1 x_1 + c_2 x_2)$$

subject to

$$\begin{aligned} \bar{P}(b_1 \leq x_1) &\leq \bar{a}_1, \\ \bar{P}(b_2 \leq x_1 + x_2) &\leq \bar{a}_2, \\ x_1 + x_2 &\leq 0, \end{aligned} \tag{20}$$

where c_1 is uniformly distributed over [15, 25], and c_2 is uniformly distributed over [$c_1 - 6$, $c_1 + 4$]. Thus the mean of c_2 , given c_1 , is $c_1 - 1$; but c_2 , like c_1 , has a range of ten units. b_1 is uniformly distributed over [100, 200] and b_2 is uniformly distributed over [75, $2b_1 - 75$]. Thus the mean of b_2 , given b_1 , equals b_1 , but the range of b_2 is not the same as the range of b_1 .

In addition it is assumed that c_1 and c_2 are independent of b_1 and b_2 . The solution of problem 20 is begun by computing $\bar{F}_1^{-1}(1-\bar{a}_1)$ and $\bar{F}_2^{-1}(1-\bar{a}_2)$.

Let y be a random variable that is uniformly distributed over [a , b]. Let \hat{y} be the $1-\bar{a}$ th fractile point of y . Then $\frac{\hat{y}}{a} = 1 - \bar{a}$. Hence $\hat{y} = (b-a)(1-\bar{a}) + a$. Thus the equivalent form of problem 20 is:

maximize

$$E(c_1 x_1 + c_2 x_2)$$

subject to

$$\begin{aligned} x_1 &\leq 100(1-\bar{a}_1) + 100 \\ x_1 + x_2 &\leq (2b_1 - 150)(1-\bar{a}_2) + 75, \\ x_1 + x_2 &\leq 0. \end{aligned} \tag{21}$$

Or, equivalently,

maximize

$$E(c_1 x_1 + c_2 x_2)$$

subject to

$$\begin{aligned} x_1 &\leq 200 - 100\bar{a}_1, \\ x_1 + x_2 &\leq 2(1-\bar{a}_2)b_1 + 150\bar{a}_2 - 75, \\ x_1 + x_2 &\leq 0. \end{aligned}$$

Let (b_1, b_2, c_1, c_2) be any sample point and let x_1 be given. Then to determine x_2^* the following must be solved:

maximize

$$c_2 x_2$$

subject to

$$\begin{aligned} x_2 &\leq 2(1-\bar{a}_2)b_1 + 150\bar{a}_2 - 75 - x_1, \\ x_2 &\geq 0. \end{aligned}$$

But $c_2 > 0$ always, so that $x_2^* = 2(1-\bar{a}_2)b_1 + 150\bar{a}_2 = 75 - x_1$. To find the optimal x_1 the following must now be solved:

maximize

$$E(c_1x_1 + c_2x_2^*)$$

subject to

$$\begin{aligned} x_1 &\leq 100(1-\bar{a}_1) + 100, \\ x_1 &\geq 0. \end{aligned} \tag{22}$$

In problem 22 E means that the expected value is taken with respect to the joint distribution of b_1 , b_2 , c_1 , c_2 . But x_1 is independent of these random variables by assumption f^\dagger and c_1 and c_2 are independent of b_1 and b_2 . Hence $E(c_1x_1) = E(c_1)x_1 = E_{c_1,c_2}(c_1)x_1 = E_{c_1}[E_{c_2|c_1}(c_2)x_1] = E_{c_1}(c_1)x_1 = 20x_1$, where the subscript of the E denotes the distribution used in computing the expectation and $c_2|c_1$ means that the conditional distribution of c_2 , given c_1 , is used.

Also

$$\begin{aligned} E(c_2x_2^*) &= E\{c_2[2(1-\bar{a}_2)b_1 + 150\bar{a}_2] + 75 - x_1]\}, \\ &= 2E(c_2b_1) + 2E(c_2\bar{a}_2b_1) + 150E(c_2\bar{a}_2) + 75E(c_2) - x_1E(c_2), \end{aligned}$$

where the expectation is computed using the joint distribution of b_1 , b_2 , c_1 , c_2 . However, in order to find x_1^* it is necessary to know explicitly the expression for $E(c_2x_2^*)$ as a function of x_1 . Hence, as $E(c_2) = E_{c_1,c_2}(c_2) = E_{c_1}[E_{c_2|c_1}(c_2)] = E_{c_1}(c_1+1) = 19$, to obtain x_1 the following must be solved:

Maximize

$$20x_1 + [-19x_1 + K]$$

subject to

$$\begin{aligned} x_1 &\leq 100(1-\bar{a}_1) + 100, \\ x_1 &\geq 0, \end{aligned}$$

where $K = 2E(c_1b_1) + 2E(c_1\bar{a}_2b_1) + 150E(c_1\bar{a}_2) + 75E(c_2)$.

Therefore

$$x_1^* = 100(1-\bar{a}_1) + 100$$

and

$$x_2^* = 2(1-\bar{a}_2)b_1 + 150\bar{a}_2 + (25 + 100\bar{a}_1) -$$

Since $E^{-1}(1-\bar{a}_1)$ was linear in b_1 , x_1^* is seen to be also linear in b_1 , which agrees with Corollary 2. Also if $\bar{a}_1, \bar{a}_2 = 1$, x_1^* is piecewise constant (in fact it is a constant), as was predicted by the discussion of the linear-programming-under-uncertainty case.

[†]In fact, from the interpretation of the general problem, it is clear that X_0 , the vector of first-period decisions, will always have $E(X_0) = X_1$. This is another direct consequence of X_1 as a zero-order decision rule (see Section 6).

Example 2

For our second example the following problem is considered:

maximize

$$E(c_1 x_1 + c_2 x_2)$$

subject to

$$\begin{aligned} \bar{P}(b_1 \leq a_{11}x_1) &\geq \bar{\alpha}_1, \\ \bar{P}(b_2 \leq a_{21}x_1 + a_{22}x_2) &\geq \bar{\alpha}_2, \\ x_1 + x_2 &\leq 0, \end{aligned} \tag{23}$$

where c_1 and c_2 are defined as in the previous example; b_1 and b_2 are jointly normally distributed with means m_1 and m_2 and variances δ_1^2 and δ_2^2 , respectively, and correlation coefficient $\rho > 0$; a_{11} and a_{22} are positive constants and a_{21} is a negative constant; and c_1 and c_2 are independent of b_1 and b_2 .

From our definition of b_1 and b_2 it can be seen that their joint frequency function is given by

$$f(b_1, b_2) = \frac{1}{2\pi\delta_1\delta_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{b_1-m_1}{\delta_1}\right)^2 - 2\rho\left(\frac{b_1-m_1}{\delta_1}\right)\left(\frac{b_2-m_2}{\delta_2}\right) + \left(\frac{b_2-m_2}{\delta_2}\right)^2\right]\right\}$$

and $f(b_2 | b_1)$, the conditional frequency function of b_2 given b_1 , satisfies

$$f(b_2 | b_1) = \frac{f(b_1, b_2)}{f_1(b_1)},$$

where $f_1(b_1)$ is the marginal frequency function of b_1 . Therefore

$$f(b_2 | b_1) = \frac{1}{\delta_2\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\delta_2^2(1-\rho^2)}\left[b_2 - m_2 - \rho\frac{\delta_2}{\delta_1}(b_1 - m_1)\right]^2\right\}$$

so that b_2 , given b_1 , is normally distributed with mean $-m_2 + \rho(\delta_2/\delta_1)(b_1 - m_1)$ and variance $\delta_2^2(1-\rho^2)$.

Thus $\bar{F}_2^{-1}(1-\bar{\alpha}_2) = \delta_2\sqrt{1-\rho^2}\Phi^{-1}(1-\bar{\alpha}_2) + m_2 - \rho(\delta_2/\delta_1)m_1 + \rho(\delta_2/\delta_1)b_1$, where $\Phi(\cdot)$ is the distribution function of a normal random variable with mean 0 and variance 1. Let k_1, k_2 be defined by

$$k_2 = \delta_2\sqrt{1-\rho^2}\Phi^{-1}(1-\bar{\alpha}_2) + m_2 - \rho\frac{\delta_2}{\delta_1}m_1,$$

$$k_1 = \rho\frac{\delta_2}{\delta_1} - 0.$$

Then a problem equivalent to problem 23 is

maximize

$$E(c_1 x_1 + c_2 x_2)$$

subject to

$$\begin{aligned} a_{11}x_1 &\leq -\delta_1\Phi^{-1}(1-\bar{\alpha}_1) + m_1 \\ a_{21}x_1 + a_{22}x_2 &\leq k_1b_1 + k_2 \\ x_1 + x_2 &\leq 0 \end{aligned} \tag{24}$$

Let (b_1, b_2, c_1, c_2) be any sample point in the space of random variables b_1, b_2, c_1, c_2 . Let x_1 be a given first-period decision rule. Then to find x_2^* the following must be solved:

maximize $c_2 x_2$

subject to

$$a_{22} x_2 \leq k_1 b_1 + k_2 - a_{21} x_1,$$

$$x_2 \geq 0,$$

or, equivalently,

maximize $c_2 x_2$

subject to

$$x_2 \leq \frac{k_1}{a_{22}} b_1 + \frac{k_2}{a_{22}} - \frac{a_{21}}{a_{22}} x_1,$$

$$x_2 \geq 0$$

as $a_{22} > 0$ by hypothesis. Since $c_2 > 0$, also by assumption, the following is obtained:

$$x_2^* = \frac{k_1}{a_{22}} b_1 + \frac{k_2}{a_{22}} - \frac{a_{21}}{a_{22}} x_1 \text{ if } k_1 b_1 + k_2 - a_{21} x_1 \geq 0,$$

no feasible solution if $k_1 b_1 + k_2 - a_{21} x_1 < 0$.

Thus region \bar{D}_1 discussed above is seen to be defined by $k_1 b_1 + k_2 - a_{21} x_1 \leq 0$. Moreover since $a_{11} > 0$ it can be seen from problem 24 that \bar{D}_1 is the region for which $\delta_1 \Phi^{-1}(1-\bar{\alpha}_1) + m_1 < 0$, so that the union of $D_1 D_2$ gives the set $\bar{\Delta}_2$.

Using the rule $x_2^* = 0$ in $\bar{\Delta}_2$, the optimal x_1 is found by solving the following:

maximize

$$E_1(c_1 x_1) + E_2(c_2 x_2^*)$$

subject to

$$a_{11} x_1 \leq \delta_1 \Phi^{-1}(1-\bar{\alpha}_1) + m_1, \quad (25)$$

$$x_1 \geq 0.$$

If $\delta_1 \Phi^{-1}(1-\bar{\alpha}_1) + m_1 > 0$, then a feasible x_1 exists and

$$E_2(c_2 x_2^*) = E_{c_2} \left(E_{c_1} (c_1 x_1) \Big|_{\bar{D}_2} \right) = \left(\frac{k_1}{a_{22}} b_1 + \frac{a_{21}}{a_{22}} x_1 \right) dF_1(x_1) + F \left(\frac{c_2 k_2}{a_{22}} \right)$$

But \bar{D}_2 is the region where $k_1 b_1 + k_2 - a_{21} x_1 > 0$, that is, $b_1 > (a_{21}/k_1) x_1 - (k_2/k_1)$ as $k_1 > 0$ and k_1 is a constant because $\bar{\alpha}_1$ is a constant (see Section 6). If it is now assumed that k_1 is also a constant, i.e., that $\bar{\alpha}_1$ is a constant, the following is obtained:

$$\begin{aligned}
E_2(c_2 x_2^*) &= 19 \int_{\frac{a_{21}x_1 - k_2 - m_1 k_1}{k_1}}^{+\infty} \left(\frac{k_1}{a_{22}} b_1 + \frac{k_2}{a_{22}} - \frac{a_{21}}{a_{22}} x_1 \right) \frac{1}{\delta_1 \sqrt{2\pi}} e^{-\frac{(b_1 - m_1)^2}{2}} db_1 \\
&= 19 \int_{\frac{a_{21}x_1 - k_2 - m_1 k_1}{k_1}}^{+\infty} \left(\frac{k_1 \delta_1}{a_{22}} z + \frac{k_1}{a_{22}} m_1 + \frac{k_2}{a_{22}} - \frac{a_{21}}{a_{22}} x_1 \right) e^{-\frac{z^2}{2}} dz.
\end{aligned}$$

Finally, using the fact that x_1 is to be independent of b_1 and c_1 , problem 25 can be written:

maximize

$$\begin{aligned}
20x_1 - \frac{19a_{21}x_1}{a_{22}} \left[1 - \Phi\left(\frac{a_{21}x_1 - k_2 - m_1 k_1}{k_1 \delta_1}\right) \right] + 19 \left(\frac{k_1 m_1}{a_{22}} + \frac{k_2}{a_{22}} \right) \left[1 - \Phi\left(\frac{a_{21}x_1 - k_2 - m_1 k_1}{k_1 \delta_1}\right) \right] \\
+ \frac{19k_1 \delta_1}{a_{22}} \int_{\frac{a_{21}x_1 - k_2 - m_1 k_1}{k_1 \delta_1}}^{+\infty} z \phi(z) dz
\end{aligned}$$

subject to

$$\begin{aligned}
a_{11}x_1 &\leq \delta_1 \Phi^{-1}(1-\bar{\alpha}_1) + m_1, \\
x_1 &\geq 0
\end{aligned} \tag{26}$$

where $\phi(\cdot)$ is the frequency function of a normal random variable with mean 0 and variance 1. Solving this problem (which can be done easily by applying the Lagrange-multiplier technique to problem 26), the following is obtained:

$$x_1^* = \frac{\delta_1 \Phi^{-1}(1-\bar{\alpha}_1) + m_1}{a_{11}}.$$

Thus the optimal decision rules are

$$x_1 = \begin{cases} \frac{\delta_1 \Phi^{-1}(1-\bar{\alpha}_1) + m_1}{a_{11}}, & \text{if } \delta_1 \Phi^{-1}(1-\bar{\alpha}_1) + m_1 \geq 0 \\ 0, \text{ i.e., no feasible } x_1 \text{ otherwise.} \end{cases}$$

$$x_2 = \begin{cases} \frac{k_1}{a_{22}} b_1 + \frac{k_2}{a_{22}} - \frac{a_{21}}{a_{22}} x_1, & \text{if } b_1 \geq \frac{a_{21}}{k_1} x_1 - \frac{k_2}{k_1} \text{ and } x_1 \text{ is feasible} \\ 0, \text{ i.e., no feasible } x_2 \text{ otherwise.} \end{cases}$$

REFERENCES

REFERENCES CITED

1. A. S. Manne, "Linear Programming and Sequential Decisions," Mgt. Sci., 6 (3): 259-67 (1960).
2. A. Charnes and W. W. Cooper, "Chance-Constrained Programming," Mgt. Sci., 6 (1): 73-79 (1959).
3. G. Tintner, "Stochastic Linear Programming with Applications to Agricultural Economics," in H. A. Antosiewicz, (ed), Second Symposium on Linear Programming, Natl. Bur. of Standards, Washington, D. C., 1955.
4. A. Charnes, W. W. Cooper, and G. H. Symonds, "Cost Horizons and Certainty Equivalents: An Approach to Stochastic Programming of Heating Oil Production," Mgt. Sci., 4 (3): 235-63 (1958).
5. ———, and M. Kirby, "Optimal Decision Rules for the Triangular Chance-Constrained Programming Problem," Research Analysis Corporation, in preparation.
6. ———, W. W. Cooper, J. K. Devoe, and D. B. Learner, "DEMON: Decision Mapping via Optimum Go-No Networks—A Model for Marketing New Products," Batten, Barton, Durstine, and Osborn, Inc., New York, Jul 63.
7. ———, and S. Thore, "Planning for Liquidity in Savings and Loans Associations," ONR Research Memo. 95, Northwestern University, Evanston, Ill., May 64.
8. ———, and M. Kirby, "Optimal Decision Rules for the Two-Period Savings and Loan Association Problem," Research Analysis Corporation, RAC-TP-174, in press.
9. ———, W. W. Cooper, and G. L. Thompson, "Critical Path Analysis via Chance-Constrained and Stochastic Programming," Opsns. Res., 12 (3): 460-70 (1964).
10. ———, ———, "Deterministic Equivalents for Different Objectives in Chance-Constrained Programming," ONR Research Memo. 37, Northwestern University, Evanston, Ill., and Carnegie Institute of Technology, Pittsburgh, Dec 60.
11. ———, ———, "Deterministic Equivalents for Optimizing and Satisficing under Chance Constraints," Opsns. Res., II (1): 18-39 (1963).
12. H. A. Simon, Models of Man, John Wiley & Sons, Inc., New York, 1957.
13. J. G. March and H. A. Simon, Organizations, John Wiley & Sons, Inc., New York 1958.
14. W. W. Cooper and J. D. Savvas, "Motivational Cost and Transients in Budgeting the Behavior of Cost and Aspirations," Behavioral Sci., in press.
15. A. Stedry, Budget Control and Cost Behavior, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1960.
16. ———, and A. Charnes, "Exploratory Models in the Theory of Budgetary Control," ONR Research Memo. 43, Northwestern University, Evanston, Ill., Jan. 62.
17. A. Ben-Israel, "On Some Problems of Mathematical Programming," Ph.D. thesis in Engineering Science, Northwestern University, Evanston, Ill., Jun 62.
18. ———, and A. Charnes, "Constant Level Inventory Policies and Chance-Constrained Programming Models," Northwestern University, Evanston, Ill., 1961.
19. G. H. Symonds, "Stochastic Scheduling by the Horizon Method," Mgt. Sci., 8 (2): 138-67 (1962).
20. G. B. Dantzig, "Linear Programming under Uncertainty," Mgt. Sci., 1 (3-4): 197-206 (1955).

21. A. Charnes, W. W. Cooper, and G. L. Thompson, "Constrained Generalized Medians and Linear Programming under Uncertainty," Mgt. Sci., in press.
22. A. Madansky, "Inequalities for Stochastic Linear Programming Problems," Mgt. Sci., 6 (2): 197-204 (1960).
23. ———, "Methods of Solution of Linear Programs under Uncertainty," Opsns. Res., 10 (4): 463-71 (1962).
24. B. V. Gnedenko, The Theory of Probability, Chelsea Publishing Co., New York, 1962.
25. A. Charnes and M. Kirby, "A Linear Programming Application of a Left Inverse of a Basis Matrix," Northwestern University, Evanston, Ill., 1963.
26. C. R. Rao, "A Note on a Generalized Inverse of a Matrix with Applications to Problems in Mathematical Statistics," J. Roy. Stat. Soc., Series B, 24: 152-58 (1962).
27. A. Charnes and W. W. Cooper, "Chance-Constrained Programs with Normal Deviates and Linear Decision Rules," Nav. Res. Logist. Quart., 7 (4): 533-44 (1960).
28. ———, ———, "Normal Deviates and Chance Constraints," J. Amer. Stat. Assoc., 57 (297): 134-48 (1962).
29. R. Penrose, "A Generalized Inverse for Matrices," Proc. Cambridge Phil. Soc., 51: 406-13 (1955).
30. A. Ben-Israel and A. Charnes, "Contributions to the Theory of Generalized Inverses," J. Soc. Indus. Appl. Math., 11 (3): 667-99 (1963).
31. G. B. Dantzig, "Recent Advances in Linear Programming," Mgt. Sci., 2 (2): 131-44 (1956).
32. ———, Linear Programming and Extensions, Princeton University Press, Princeton, N.J., 1963.
33. ———, and A. Madansky, "On the Solution of Two-Stage Linear Programs under Uncertainty," Proc. Fourth Berkeley Symposium Math. Stat., 1961.
34. A. R. Ferguson and G. B. Dantzig, "The Allocation of Aircraft to Routes—An example of Linear Programming under Uncertain Demand," Mgt. Sci., 3 (1): 45-73 (1956).

ADDITIONAL REFERENCES

- A. Charnes, and W. W. Cooper, "The Theory of Search: Optimum Distribution of Search Effort," Mgt. Sci., 5 (1): 44-50 (1958).
- , ———, Management Models and Industrial Applications of Linear Programming, John Wiley & Sons, Inc., New York, 1961, Vols 1 and 2.
- Elmaghraby, S. E., "An Approach to Linear Programming under Uncertainty," Opsns. Res., 7: 208-16 (1959).
- Tintner, G., "A Note on Stochastic Linear Programming," Econometrica, 28 (2): 490-95 (1960).